

## **General Disclaimer**

### **One or more of the Following Statements may affect this Document**

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

04

S7

E77-10064

NASA CR. 149263

"Made available under NASA sponsorship  
in the interest of early and wide dis-  
semination of Earth Resources Survey  
Program Information and without liability  
for any use made thereof."

SR No. 29960

RECEIVED BY  
NASA STI FACILITY

DATE:

12/24/76

DCAF NO.

011095

PROCESSED BY

☒ NASA STI FACILITY☐ ESA - SDS ☐ AIAA

AGRICULTURE/FORESTRY  
HYDROLOGY

Mr. W.J. van der Oord *auto*  
Mekong Secretariat  
c/o ESCAP Sala Santitham  
Bangkok, Thailand

November 1976

(E77-10064) AGRICULTURE/FORESTRY HYDROLOGY  
Quarterly Report, Sep. - Nov. 1976 (Mekong  
Committee Secretariat, Bangkok) 40 p  
HC A03/MF A01

N77-14563

CSCL 08H

G3/43

Unclass

00064

Type II Quarterly Report

Mr. Frederick Gordon  
Technical Monitor  
Code 902  
NASA/Goddard Space Flight Center  
Greenbelt, Maryland 20771

29960

RECEIVED

NOV 29 1976

SIS/902.6

- |   |   |                            |
|---|---|----------------------------|
| 1. SR No.   | 2. Type of Report   | 3. Recipient's Catalog No. |
| 29960   | II  |                            |
| 4. Title  | 5. Report Date  |                            |
| Agriculture/Forestry<br>Hydrology   | 10 November 1976  |                            |
|   | 6. Period Covered   |                            |
|   | September 1976-November 1976  |                            |
| 7. Principal Investigator   | 8. No. of Pages :   | 40                         |
| W.J. van der Cord   |   |                            |
| 9. Name and address of Principal Investigator's Organisation                                      | 10. Principal Investigator Report No. 5   |                            |
| Mekong Secretariat<br>c/o ESCAP, Sala Santitham<br>Bangkok, Thailand                              |   |                            |
|   | 11. CSFC Technical Monitor  |                            |
|   | Mr. Frederick Gordon  |                            |
| 12. Sponsoring Agency Name and Address  | 13. Key Words (Selected by Principal Investigator)                              |                            |
| Mr. Frederick Gordon<br>Code 902<br>NASA/Goddard Space Flight Center<br>Greenbelt, Maryland 20771 | - Color diazochrome films<br>- Mapping programme<br>- Computer Compatible Tapes |                            |
| 14. Supplementary Notes   |   |                            |
| 15. Abstract  |   |                            |

## TYPE II QUARTERLY REPORT

### I. INTRODUCTION

The main objectives of the Mekong Committee investigations using LANDSAT-2 data are as follows :

#### A. Short-term objectives

By both photo-interpretation and automatic data processing techniques, supported by ground truth data and field surveys, establish :

- Land use, land capability and hydro-geomorphology maps of the lower Mekong basin;
- Maps showing primary forests and deciduous forest areas;
- Maps showing consecutive flood and drainage patterns of Mekong lowlands.

#### B. Long-term objectives

Organise a research programme for classification of agricultural crops and land use, and for soil moisture monitoring for crop forecasting.

### II. TECHNIQUES

#### 1. Data quality and delivery

The data received during the period under review is listed in Annex 1.

All scenes are outside of our investigation site. In this connexion, it should be noted that NASA had refunded our LANDSAT Account No. G29960 of an amount of US\$4,264 equivalent to the cost of imagery previously delivered but outside of our investigation site due to computer error in search. As a result, the balance of our LANDSAT Account was US\$5,176, as of 31 October 1976.

As for the delivery of data, by letter of 24 September 1976 NASA informed that the period of data acquisition for our standard products terminates on 1 January 1976, with possibility to order data retroactively against our account at Eros Data Center. Following this instruction an order of imagery over twenty scenes had been placed.

2. Preparation of land use, land capability and morpho-pedologic maps

As part of the general thematic mapping programme, two computer programmes for processing data on CCT were developed by one Mekong Secretariat staff trainee at the Grounement pour le Développement de la Télédétection Aérospatiale (G.D.T.A.) in France.

The DETECT programme makes/<sup>it</sup> possible/<sup>to</sup> delineate the boundary between two areas having low contrast. The SPACLA programme determines homogeneous areas and identifies them in relation to selected classes. A description of these two programmes is provided in Annex 2. The method is specially used in the case of areas for which ground truth control is practically unavailable.

3. Research programme for computer processing of remote sensing data

For the benefit of information interpretation a programme for computer processing of ground truth data was developed. Determination of various cropping feature which can be differentiated and identified must be condensed in the way that computer can accept. All data must be converted to the computer accessible form before the compilation and storing steps.

Seven main features were collected, i.e. rice, grass, bush, housing or factory, bareground, water and soil conditions. During the field observation, the environmental and atmospheric conditions were also recorded.

- 3 -

The data was checked and edited both manually and by a computer processing before storing in computer files for fast retrieval. To retrieve the information from the file, observation date and test site type are selected as the primary indices for each data set. Then each record may be retrieved in terms of its field identification.

The computer processing of ground truth data is described in detail in Annex 3.

### III. ACCOMPLISHMENTS DURING THE PERIOD UNDER REVIEW

#### 1. Preparation of land use, land capability and morpho-pedologic maps

The preparation of the final map of the land use map at the scale of 1:1,000,000 was completed. At the present time, the map is being published.

Work continued on the preparation of the land capability and morpho-pedologic maps.

#### 2. Research programme for computer processing of remote sensing data

##### 2.1 Ground truth data collection

Two basic data collection systems for obtaining the ground truth observations during the time that LANDSAT-2 satellite passes over two selected sites for a period of one year starting from August 1975 have been completed. These test sites are the area along Paholyothin Highway from Km. 29 just north of Don Muang Airport to Km. 65 (Wang Noi District) and the area in Tambon Bang Khan, a subdistrict of Klong Luang District, Pratumthani Province. Ground information collected indicate cropped changing stages, and the man-made objects along the Highway test site permit the ground check points to identify and to map against the desired satellite imagerial pixels.

The ground truth data was checked and edited both manually and by a computer processing before storing in computer files for fast retrieval, as mentioned earlier in section II - Techniques.

## 2.2 Computer processing of remote sensing data

The following computer programmes development have been completed during the period under review.

- Six programming phases of the RECOG package have been modified and installed on the IBM 370/145 computer system. Brief description of the six programming phases is provided in Annex 4.

- Sequential input-output files of the package have been reorganized to suit a random processing procedure. This substantially reduces input-output searching time.

- The cubic convolution technique was added to the package. The technique provides smooth computer print-out images and will help users in extraction of information in small sub-image areas where individual pixels can be seen.

- Test run using LANDSAT-1 computer compatible tape of frame No. 5 - 6 was to classify the image around Wang Noi District. Eleven sampling sets have been processed by the modified RECOG package.

## IV. CONCLUSIONS

Work continued as planned on the general mapping programme. Substantial progress have been made on the research programme for computer processing of remote sensing data following the completion of the ground truth data collection and the installation of the RECOG package on IBM 370/145.

\*\*\*\*\*

ORIGINAL PAGE IS  
POOR QUALITY

ANNEX 1

LIST OF LANDSAT-2 IMAGERY

RECEIVED BY THE MEKONG SECRETARIAT

(Period September-November 1976)

ID number	Cover the Mekong basin	Outside the Mekong basin	Date acquired (month,day,year)	Remarks (localization)
1 2368-03305		x	01/25/76	Burma
2 2368-03312		x	01/25/76	Bay of Bengal
3 2368-03314		x	01/25/76	Bay of Bengal
4 2368-03321		x	01/25/76	Bay of Bengal
5 2370-08521		<u>x</u>	01/27/76	Niger
Total		5		

ANNEX 2

DESCRIPTION OF  
THE DETECT AND SPACLA PROGRAMMES

DETECT PROGRAMME

DETECTION AND TRACING OF INTER-ZONAL  
BOUNDARIES ON MULTISPECTRAL IMAGES

---

BY

NGUYEN DUONG

# 1. THE PROBLEM

Several approaches have already been developed to detect and trace the interzonal boundaries of images given by the multispectral scanner; these techniques however are either too complex and time consuming (as for instance techniques using the FOURIER, HADAMARD or WALSH transformed curve), or too simple and insufficient for a proper determination of interzonal boundaries when data are marred by noises or when contrasts between neighbouring areas are too weak.

This report describes a simple but efficient programme - named DETECT and written in FORTRAN - which makes it possible to detect and trace boundaries from multispectral images which have not been substantially affected by noises.

It can be used to delimit homogeneous areas in the zone to be studied and will therefore help in computing correctly the class statistics of objects located in this zone.

REPRODUCIBILITY OF THE  
ORIGINAL PAGE IS POOR

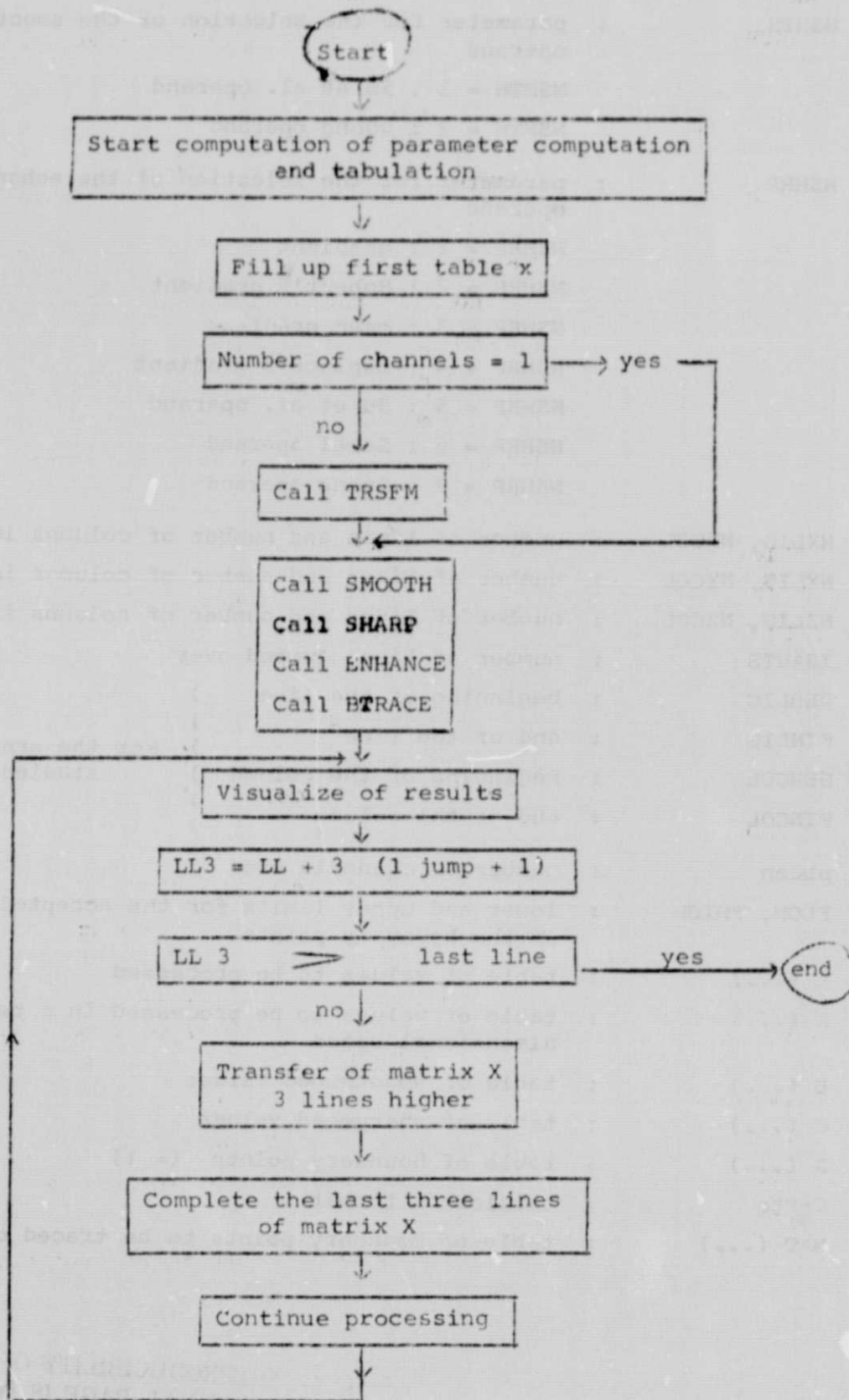
## 2. GENERAL ORGANIZATION OF THE DETECT PROGRAMME

This programme is composed of five successive operations aimed at accentuating contrasts (tracing boundaries and visualizing these borders on film). These five operations are carried out through sub-programmes respectively named SMOOTH, SHARP, ENHANCE, BTRACE and BVISU.

Data used as input for the DETECT programme appear as a table with an odd number of lines higher than five (seven for instance).

The general organization of the DETECT programme is given in the following diagram.

DETECT PROGRAMME



ORIGINAL PAGE IS  
OF POOR QUALITY

MEANING OF PARAMETERS AND TABLES IN THE MAIN PROGRAMME :

NMOV : size of the smoothing operand

NSMTH : parameter for the selection of the smoothing operand  
 NSMTH = 1 : Su et al. operand  
 NSMTH = 2 : Duong operand

NSHRP : parameter for the selection of the enhancing operand  
 NSHRP = 1 : gradient  
 NSHRP = 2 : Robert's gradient  
 NSHRP = 3 : mean gradient  
 NSHRP = 4 : Laplace's gradient  
 NSHRP = 5 : Su et al. operand  
 NSHRP = 6 : Sobel operand  
 NSHRP = 7 : Duong operand

NXLIG, NXCOL : number of lines and number of columns in table X

NYLIG, NYCOL : number of lines and number of columns in table Y

NZLIG, NZCOL : number of lines and number of columns in table Z

ISAUTE : number of lines jumped over

DEBLIG : beginning of the line )

FINLIG : end of the line )

DEBCOL : beginning of the column ) For the area to be studied

FINCOL : end of the column )

nbcan : number of channels used

FLOW, FHIGH : lower and upper limits for the accepted values of the boundary points

X (...) : table of values to be processed

A (...) : table of values to be processed in a two dimensional space

B (...) : table of smoothened values

C (...) : table of sharpened values

D (...) : table of boundary points (= 1)

Carte : tabulation format

MAT (...) : table of boundary points to be traced on film.

### 3. SUB-PROGRAMMES

#### 3.1 SMOOTH :

This is a smoothing programme intended to cancel noises and reduce sudden variations.

Two smoothing operands are used in the SMOOTH programme:

- The Su et al. operand : a table  $(2N + 1) \times (2N + 1)$  determined by

$$Y(i, j, k) = \frac{1}{(2N + 1)^2} \sum_{i'=-N}^N \sum_{j'=-N}^N X(i + i', j + j', k)$$

where  $y(i, j, k)$  is the smoothened value of the initial value of point  $x$  having  $(i, j)$  as co-ordinates in the  $k$  spectral band.

- The operand proposed by the author :  $3 \times 3$ , which is found in the following manner :

assuming NBH (m) with  $m = 1; 2 \dots 8$ , corresponding to the eight points located in the immediate neighbourhood of point  $x(i, j, k)$ , and  $\sigma$ , the standard deviation in relation to these points. Points retained for the smoothing operation will have to be located close to point  $x(i, j, k)$  and to satisfy to the following condition

$$|x(i, j, k) - \text{NBH}(m)| \leq a \sigma$$

where  $a$  is a multiplication factor selected by the user. If none of the eight points located in the immediate neighbourhood of point  $x(i, j, k)$  satisfies to the above condition, the average of neighbouring points will be retained as the smoothened value of point  $x(i, j, k)$ .

This algorith will eliminate shot noises and, when used several times (three times for instance), will give quite homogeneous areas with clear boundaries.

Meaning of parameters and tables in the SMOOTH sub-programmes

NSMTH : parameter for the selection of the smoothing operand  
 NMOV : size of the smoothing operand  
 A (...) : table of values to be processed in a two-dimensional space  
 B (...) : table of smoothened values  
 OP ( . ) : table of the smoothing operand  
 PNBH ( . ) : eight points located in the immediate neighbourhood of point P

3.2 SHARP :

The purpose of this programme is to sharpen boundaries.

Since smoothing (or integration) "jams" a picture, contrasts will have to be naturally improved through a differentiation operation.

The following operands are introduced in the SHARP programme :

- Robert's gradient

$$|\nabla P|_R^2 \triangleq \left[ P(i+1, j+1) - P(i, j) \right]^2 + \left[ P(i+1, j) - P(i, j+1) \right]^2$$

- Mean gradient

$$|\nabla P|^2 \triangleq \frac{1}{4} \left\{ \left[ P(i, j) + P(i, j+1) - P(i+1, j) - P(i+1, j+1) \right]^2 + \left[ P(i, j) + P(i+1, j) - P(i, j+1) - P(i+1, j+1) \right]^2 \right\}$$

- Su and al. operand

$$|\nabla P|_{Su}^2 \triangleq \left[ P(i, j) - P(i, j-d) \right]^2 + \left[ P(i, j) - P(i, j+d) \right]^2 + \left[ P(i, j) - P(i, d, j) \right]^2 + \left[ P(i, j) - P(i, d, j) \right]^2$$

- Sobel operand :

$$Z_1 = P(i-1, j-1) + 2P(i, j-1) + P(i+1, j-1)$$

$$Z_2 = P(i-1, j+1) + 2P(i, j+1) + P(i+1, j+1)$$

$$Z_3 = P(i-1, j-1) + 2P(i-1, j) + P(i-1, j+1)$$

$$Z_4 = P(i+1, j-1) + 2P(i+1, j) + P(i+1, j+1)$$

$$|\nabla P|_{So}^2 \triangleq \frac{1}{16} \{ (Z_1 - Z_2)^2 + (Z_3 - Z_4)^2 \}$$

- Operand proposed by the author :

$$Z_1 = |P(i-1, j) - P(i, j+1)| + |P(i-1, j+1) - P(i, j)|$$

$$Z_2 = |P(i, j+1) - P(i+1, j)| + |P(i+1, j+1) - P(i, j)|$$

$$Z_3 = |P(i+1, j) - P(i, j-1)| + |P(i+1, j-1) - P(i, j)|$$

$$Z_4 = |P(i, j-1) - P(i-1, j)| + |P(i-1, j-1) - P(i, j)|$$

$$|\nabla P|_D \triangleq \frac{1}{4} (Z_1 + Z_2 + Z_3 + Z_4)$$

Meaning of parameters and tables in the SHARP sub-programme :

NSHRP : parameter for selection of the sharpening operand

B ( ... ) : table of smoothened values

C ( ... ) : table of sharpened values.

3.3 ENHANCE :

The purpose of this programme is to select automatically points with the highest probability of being boundary points.

Assuming  $\bar{P}(k)$  and  $\bar{S}(k)$  as corresponding to the mean value and the standard deviation of points  $P(k)$  in channel  $k$ ; that  $T(k)$  is the threshold retained for the automatic selection of appropriate points in channel  $k$  for the sub-programme; if  $P(i,j,k)$  is greater or equal to  $T(k)$ , in one of the channels

at least, it is considered that this point is located on or near the boundary sought.

Su and al. have recommended that the value of  $T(k)$  be computed on the basis of the following equation :

$$T(k) = \bar{P}(k) + a \bar{S}(k)$$

where  $a$  is a multiplication factor selected by the user.

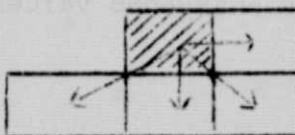
#### Meaning of parameters and tables in the ENHANCE sub-programme

- FLOW, FHIGH** : lower and higher limits for accepted values of boundary points
- DELT** : threshold for acceptable variations
- C ( ... )** : table of sharpened values
- D ( ... )** : table of border points ( = 1)

#### 3.4 BTRACE :

This programme is used for the detection of boundaries. Its purpose is to select boundary points among possible points and to decide of the direction(s) to be followed to trace the boundaries.

The detection of boundary direction is done through a 2 x 3 operand, which can select among four possible directions to trace the boundary(ies) as below :



Operand for boundary detection

This operand has been selected so as to be compatible with images supplied by a multispectral scanner, input data for the BTRACE programme are read line after line and points tested from the left to the right.

/ Meaning

### Meaning of parameters and tables in the BTRACE sub-programme

DELT : threshold for acceptable variations  
 C ( ... ) : table of sharpened values  
 D ( ... ) : table of boundary points ( = 1 )  
 ZNBH ( . ) : four points of the operand located in the immediate vicinity of the point considered.

### 3.5 BVISU :

This programme sees to the visualization of boundaries on film. It relies necessarily on the CALCMP sub-programme of the Centre National d'Etudes Scientifiques library of programmes to trace curves linking boundary points.

### Meaning of parameters and tables in the BVISU sub-programme

NN : number of the line to be traced  
 NM : number of points per line  
 CREM : increment of film shift  
 BOUND ( ... ) : table of boundary points to be traced  
 NBH ( . ) : four points of the operand which are located in the immediate vicinity of the point considered.

### 3.6 TRSFM :

The purpose of this programme is to transform the multi-spectral image in a two-dimensional image. It makes use of distances to replace point co-ordinates within the "boundary" space.

### Meaning of parameters and tables in the TRSFM sub-programme

NBCAN : number of channels used  
 X ( ..... ) : initial table of values to be processed  
 A ( ... ) : table of transformed values  
 P ( . ) : co-ordinates of the point considered in the "boundary" space.

#### 4. RESULTS

The DETECT programme has been tested with ERTS data on the Mekong basin and with AIRBORNE data on the Bouches-du-Rhône (Rhône delta).

##### 4.1 ERTS data on the Mekong basin

Since channel 6 is largely marred by noise, the only data used to test the DETECT programme were those of channels 4, 5 and 7.

Values selected for the NMOV, NSMTH, NSHRP, FLOW and FHIGH parameters are the following :

NMOV	=	3
NSMTH	=	2
NSHRP	=	5
FLOW	=	0.2
FHIGH	=	0.6

These were tested with data corresponding to an area of strip 1 limited by

lines, 1800 to 2100  
and columns, 200 to 500.

The only boundaries that appear clearly on the chart correspond to the reservoir located near Ban Khok Sung, the Nam Phong canal and the cultivated area north of the reservoir; the rests are only separate parts of curves. This can easily be explained, since with the exception of the cultivated area, no wide enough homogeneous areas to correspond to a specific type of object, the only thing found is mixed open forests with few crops.

##### 4.2 AIRBORNE data on the Bouches-du-Rhône (Rhône delta)

Only channels 5, 7 and 10 were retained to test the DETECT programme. The values retained for parameters NMOV, NSMTH, NSHRP, FLOW and FHIGH are the same as before.

/ These

These are tested with data corresponding to an area limited by :

line, 600 to 960 (jumped over 1)  
columns, 60 to 310

Since AIRBORNE data are less marred by noise than ERTS data, and since the resolution elements (IFOV) is much smaller, practically all boundaries in the area considered appear clearly on the chart.

## 5. CONCLUSION

After testing the DETECT programme with ERTS data on the Mekong basin and AIRBORNE data on the Bouches-du-Rhône, we can make the following comments :

5.1 The DETECT programme is efficient in tracing boundaries of areas that are homogeneous or are not marred by noise. In the later case smoothing operations must necessarily cover a larger number of points (5 x 5 or 7 x 7 for instance).

5.2 SMOOTH and SHARP sub-programmes can be used to enhance the texture of the area considered, which is most useful for classification programmes.

5.3 Sub-programmes are unrelated to each other and operands used to smooth or sharpen data are left to the users' discretion, the DETECT programme thus has a greater flexibility and can be used with different data.

5.4 The DETECT programme has been prepared rather quickly; an improvement of this programme is recommended both regards the programming method and the presentation of results.

SPACLE PROGRAMME

CLASSIFICATION THROUGH SPATIAL CLUSTERING

---

BY

NGUYEN DUONG

# 1. THE PROBLEM

The techniques of classification through clustering are natural approaches used to group similar objects within an area. The main difficulties in applying the techniques relate to computation complexities and to the size of tables used in the classification programmes. So as to better adapt these factors to multispectral imagery data, NAGY proposed a line by line sequential analysis method and a classification by clusters; however, since reflection characteristics of objects belonging to a same cluster can change from one line to another because of atmospheric perturbations, neighbouring points tend in fact to belong to a same cluster, with the exception of those located on the boundaries. This concept brought about the development of a programme for classification by spatial clustering written in FORTRAN and known as SPACLA. It stands as an intermediary between the controlled classification programme and the uncontrolled classification programme (CLANS) which are already available in the programme library of the C.N.E.S. (Centre national d'études scientifiques). This programme which can be used to classify the Mekong data according to the various classes selected earlier, is one of the main tools for the preparation of thematical maps relating to this area.

REPRODUCIBILITY OF THE  
ORIGINAL PAGE IS POOR

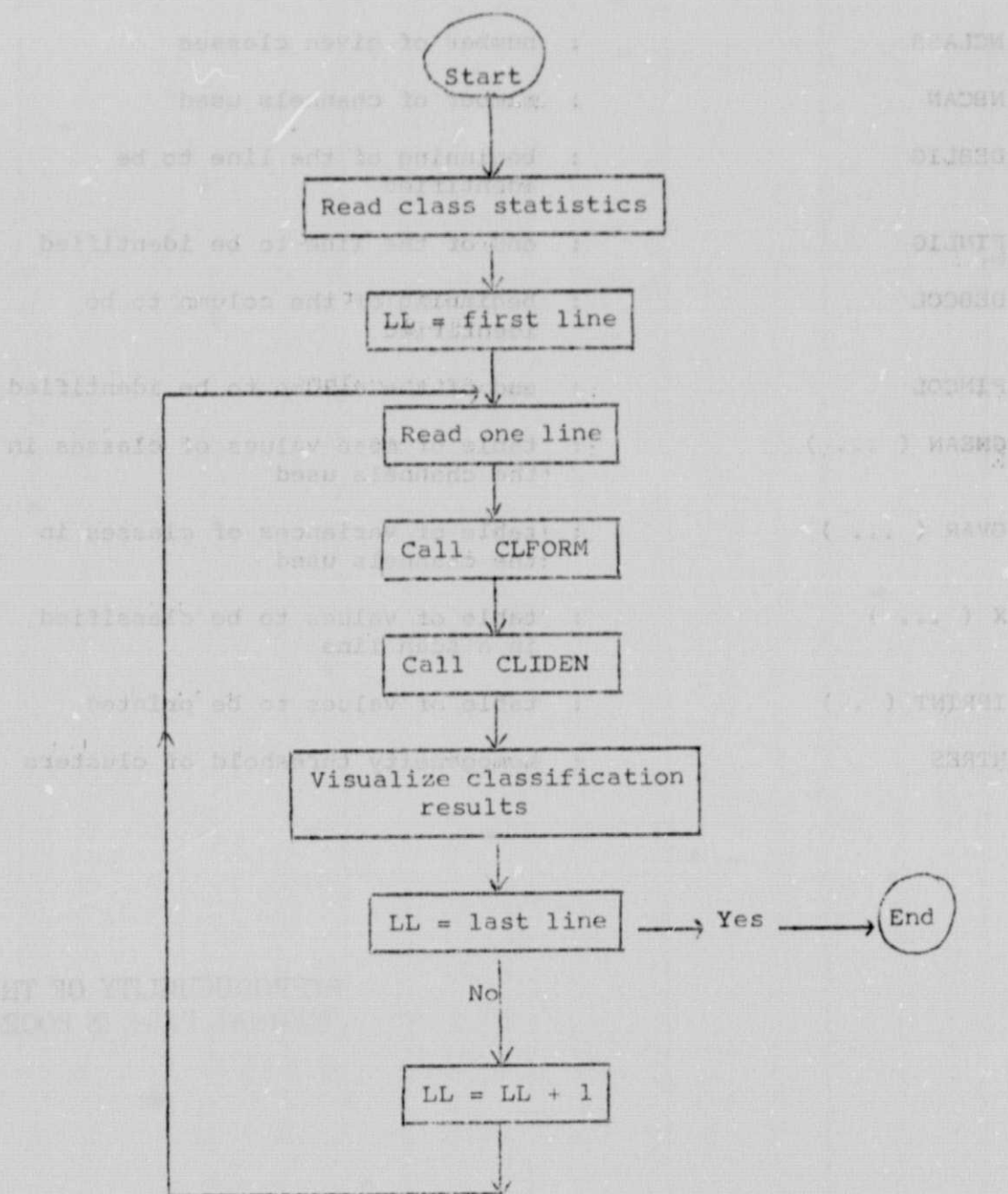
## 2. OVERALL STRUCTURE OF THE SPACLA PROGRAMME

The SPACLA programme is composed of two main elements :  
a) the formation of homogeneous clusters of points for each scan line and b) the identification of these clusters in relation to predetermined classes. These two functions are respectively carried out by the CLFORM and the CLIDEN sub-programmes.

Input data for the SPACLA programme are given line by line. Classes statistics (average and variance) are obtained earlier through the SEQCLA programme and read at the beginning of the SPACLA programme routine.

The overall structure of the SPACLA programme is given in the following diagramme.

SPACLA PROGRAMME



ORIGINAL PAGE IS  
OF POOR QUALITY

MEANING OF PARAMETERS AND TABLES IN THE MAIN PROGRAMMES :

NCLASS	: number of given classes
NBCAN	: number of channels used
DEBLIG	: beginning of the line to be identified
FINLIG	: end of the line to be identified
DEBCOL	: beginning of the column to be identified
FINCOL	: end of the column to be identified
GMEAN ( ... )	: table of mean values of classes in the channels used
GVAR ( ... )	: table of variances of classes in the channels used
X ( ... )	: table of values to be classified in a scan line
IPRINT ( . )	: table of values to be printed
HTRES	: homogeneity threshold of clusters

REPRODUCIBILITY OF THE  
ORIGINAL PAGE IS POOR

### 3. SUB-PROGRAMMES

#### 3.1 SPCLUS

The purpose of this programme is to form homogeneous clusters of points on each scan-line. Results obtained as output of this sub-programme depend to a large extent on the type of "distance measurement" selected to measure to degree of association of the points.

Since it depends on the effective closeness among couples of points studied, the association is an independent operation. Similarity, on the other hand, is determined by comparing all associations among the given points. It depends on the cluster of reference points selected to represent the environment. Therefore, two very similar points may not be closely associated, but two closely associated points are generally very similar and fully associated points are fully similar.

If  $P_{ij}$  is the association index between two points  $i$  and  $j$  ( $P_{ij}$  tends to be equal to zero when "i" and "j" are very far from each other, and to 1 when "i" and "j" are closely associated), formal definition of the similarity index between two points "i" and "j" in relation to "reference cluster"  $K$  is obtained through the following equation :

$$S_{ij} = 1 - \frac{1}{K-2} \sum_{\substack{k \in K \\ k \neq i, j}} |P_{ik} - P_{jk}|$$

$S_{ij}$  will be equal to zero when points "i" and "j" are fully unsimilar and nearly equal to 1 when these 2 points are very similar. On the basis of the association and similarity concepts,  $d_{ij}$ , the "resemblance index", can be determined as follows :

$$d_{ij} = \frac{2S_{ij} P_{ij}}{S_{ij} + P_{ij}}$$

In the SPCLUS programme, the association index and the resemblance index are respectively computed by the DIST and CLIND sub-programmes.

#### MEANING OF PARAMETERS AND TABLES IN THE SPCLUS SUB-PROGRAMME

HTRES	:	homogeneity threshold of clusters
CTHRES	:	association threshold of clusters
NBCLUS	:	number of classes in a scan-line
NSIZE ( . )	:	table of clusters size
NCLASS	:	number of given classes
CLUST ( ..... )	:	temporary location of values of the elements of a same cluster
X ( ... )	:	table of values to be classified in a scan-line
CMEAN ( . )	:	table of average values of a cluster in the channels used
CVAR ( . )	:	table of variances of a cluster in the channels used.

#### 3.2 CLIDEN

This programme is used to identify homogeneous clusters in relation to given classes, the statistics being read at the beginning of the SPACLA programme routine.

The CLIDEN programme uses the CLUMP sub-programme to decide whether two clusters (one being known and the other being identified) are similar or not. The resemblance criteria of these two clusters is determined as follows :

If  $d_i$  and  $d_j$  are the distances respectively separating the centers of the  $i$  and  $j$  clusters and the surface of the dispersion ellipsoid following the direction of the line joining the two centers; and if  $D_{ij}$  is the distance between the two

/ centers,

centers, the two clusters can be considered as belonging to a same class when :

$$D_{ij} \leq a (d_i + d_j)$$

a being a multiplication parameter known as "clustering threshold", which can vary between 0.8 and 1.

#### MEANING OF PARAMETERS AND TABLES IN THE CLIDEN SUB-PROGRAMME

NCLASS	:	number of given classes
NBCAN	:	number of channels used
CTHRES	:	association threshold of the clusters
GMEAN ( ... )	:	table of average values of classes in the channels used
GVAR ( ... )	:	table of variances of classes in the channels used
IGRAD ( . )	:	table of symbols to be printed.

REPRODUCIBILITY OF THE  
ORIGINAL PAGE IS POOR

#### 4. RESULTS

The SPACLA programme has been tested with ERTS data on the Mekong basin. Channels 4, 5 and 7 were the only channels used for clustering and identification of points within the areas studied. The selected values of parameters HTRES, CTHRES and NCLASS are the following :

HTRES = 0.90

CTHRES = 1

NCLASS = 7

#### 5. CONCLUSION

After testing the SPACLA classification programme with ERTS data on the Mekong basin, the following comments can be made :

5.1 Classification depends to a large extent on the value of CTHRES, the clustering threshold. If CTHRES is too small, only those clusters that are practically identical to known classes are classified. In this case, data noises appear clearly, thus giving users an idea of the quality of spectral tapes.

5.2 If CTHRES = 1 and channels selected are less marred by noises (5 and 7), satisfactory classification results can be obtained after comparison with ground truth data.

5.3 It is recommended that the application of the SPACLA programme be tested for the classification of data relating to strips 2, 3 and 4 and for their use with air-borne data on the Bouches-du-Rhône area (Rhône delta).

ANNEX 3

GROUND TRUTH DATA COLLECTION

# I. DATA COLLECTION AND DATA PREPARATION

## Ground truth observation sites and their observed conditions

Two sites were selected for the ground truth observation :

A. Site Number 1 : Along Paholyotin Highway. This site covers from Km. 29 along the highway just north of Don Muang Airport to Km. 65 (Wang Noi District).

The site was selected for possible maximum coverage with the 36 kilometre distance on the ground. Since the data collection on this site based on the distance unit basis, power electric poles and kilometre marks of the Highway were taken to identify the location of the terrain and crop conditions. The following expected conditions between the electric poles internal and/or kilometre marks were taken :

- 1) If the electric poles were present at the moment of observing data, the interval of the poles would be recorded. Otherwise, approximated half kilometre distance between two adjacent kilometre marks would be recorded.
- 2) Rice field growing stage indicated whether it was land preparation, seedling, transplanting, tillering, flowering, ripening or Harvesting condition.
- 3) Biophysical rice conditions, its height, density, and colour.
- 4) Water condition on the ground.

B. Site Number 2 : Double cultivation of the rice field at the vicinity of Tumbon Bang Khan near AIT.

This site is a fixed farm size of approximately 55 hectares (137.5 acres). The following field conditions were observed and

/recorded

ORIGINAL PAGE IS  
OF POOR QUALITY

recorded by trained observers :

- 1) Field identification indicated by its homogeneity of field condition on the observation date.
- 2) Cropping types, whether a specific field was rice, grass, bush, or bareground.
- 3) Rice field growing stage, whether it was land preparation, seedling, transplanting, tillering, flowering, ripening, or harvesting condition.
- 4) Biophysical condition, indicating the height, density, colour, and uniformity of the cropping type.
- 5) Water condition, indicating the depth, colour and clarity of water in the observing field.
- 6) Soil condition, indicating the water-holding capacity of the soil in the observing field.

In conjunction with the LANDSAT-2 satellite, these two sites selected are in the overlapped area of the satellite frame 4-5 and 5-6. This gives twice better chance of obtaining a cloud free coverage.

#### Ground truth observation and data preparation

Ground truth observation on both mentioned sites had been carried out corresponding to the schedule of the passes of LANDSAT-2 satellite. Observation and data collection were planned to be made as close as possible to the schedule except when wheather condition did not permit.

Observation and data collection were carried out by two separated teams, one along the highway (Site Number 1) and other at the Bang Khan (Site Number 2). Observation along the highway was gathered by a team of 3 persons on a vehicle running along the highway. The terrain or crop conditions on the left hand side

of the highway were observed and recorded. Photographs were also taken at every kilometre mark. Total hours per observation date took about 4 hours with the running 30 km/h . vehicle. The photographic records showing the changing stage of the rice field at Kilometre mark No. 59.

Owing to the considerable size of Site Number 2, the field was reasonably divided into three portions. The field conditions of each portion were gathered by a group of two persons. It normally took about 3-4 hours to complete the observation under the moist field condition.

In pursuit of these data collection procedures, the recent observing data must be checked and studied. As earlier mentioned, the compiling step must be completed in a short turn-around time so that the refilling data does not give any discrepancies according to the lagging time. It is also costly to re-observe the data over the whole site again. Therefore, the most likely and practical way is to prepare and instruct necessary details of the preceding collection for the observers just enough for guidance in order to obtain the least distorted information.

The following sequential data preparation steps have been executed after obtaining the data from the field :

- 1) Manual check to see if any contradiction occurs with respect to the corresponding data observed previously. If contradiction was detected, the re-observation in the field will be carried out.
- 2) Automatic check by computer will be described in the next chapter.
- 3) After the data set passes through step 1 and step 2, the field identification, area usage, and planting condition will be given to the observers for the succeeding period of both sites. Particularly, in

/ Site

Site Number 2 three separate portions of rice field map will be prepared including the present boundaries of each field identificational. Such historical information and map details will provide the observers the necessary check-points. In other words, they can easily check their locations and what they are observing.

In addition to the field data collection, the general environmental and atmospheric conditions have been recorded at AIT Water Resources Laboratory, while those two teams of observers go out to the field.

## II. COMPUTER PROCESSING OF GROUND TRUTH DATA

It is a principal focal point in developing automatic recognition processing for the benefit of information interpretation. Determination of various cropping feature which can be differentiated and identified must be condensed in the way that computer can accept. Certainly, all data must be converted to the computer accessible form before the compilation and storing steps.

### Conversion of data to computer accessible form

The significant level of accessing to the computer must be well designed so that the details of the observed data are precise and the least error in converting the information can be achieved. The basic computer accessible form for this data observation is the standard 80 column punch card. The descriptions of both ground truth observation sites are specified in a punch card format.

The common definitions of these two sites are specified in the first column through the thirty second column of the punched card. The common information contents are observation date,

cropping conditions, rice field conditions, and biophysical conditions. The specific details of Site Number 2 rice field are water condition (col. 35-40), soil condition (col. 42-47), and field identification (col. 50-54). Since the highway data observation is based on the distance unit basis, it needs two referring distance marks : The range of electric poles (col. 57-64) and the interval of kilometre marks (col. 73-80).

#### Compilation and screening of data

The computer programming has been implemented on IBM 370/145 to compile the observed data which converted to the card format. In any of the following conditions occur, the computer will print a "x" symbol aside such non-existent data :

- 1) If the observation date is different from the given observation date.
- 2) If the cropping condition code is neither the assigned value nor realistic; for instant, the condition indicates rice field and bareground simultaneously.
- 3) If the rice field indicated without rice field condition, or vice versa.
- 4) If the biophysical rice field condition is out of the given lower and upper bounds.
- 5) In case of Site Number 2 rice field area, the additional conditions are :
  - a) if soil condition indicated wet or submerged without water depth, or vice versa;
  - b) the unrealistic depth of water for area under water;
  - c) if the field identification is not a member of the given field areas and observation date.

6) In case of observation along the Highway, the additional conditions are :

- a) if the electric pole number is out of the given range;
- b) if the Kilometre mark is out of the given range.

#### Ground truth information file

In conjunction with the LANDSAT-2 satellite imagerial computer interpretation, the overall periodical observation data of both mentioned sites are stored in a fashion that any data set can easily be retrieved.

Each test site contains about 250 records per observation date and there are 18 data sets of each site. The main reason for the increasing number of records because of the different timing cultivation of the rice field.

To retrieve the information from the computer file, observation date and test site type are selected as the primary indices for each data set discrimination. Then each record will be retrieved by its field identification.

The observed data sets have been stored in six IBM diskettes in which each diskette contains 8 data sets.

### III. SUMMARY

Two basic data collection systems for obtaining the ground truth observations during the time that LANDSAT-2 satellite passes over two selected sites for a period of one year starting from August 1975 have been described. These test sites are the area along Paholyothin Highway from Km. 29 just north of Don Muang Airport to Km 65 (Wang Noi District) and the area in Tambon

Bang Khan, a subdistrict of Klong Luang District, Pratumthani Province. Ground information collected indicate cropped changing stages, and the man-made objects along the Highway test site permit the ground check points to identify and to map against the desired satellite imagerial pixels.

Seven main features in the above mentioned sites were collected, i.e., rice, grass, bush, housing or factory, bare-ground, water and soil conditions. During the field observation, the environmental and atmospheric conditions at AIT Water-Resources Laboratory were recorded. Colour photographs were also taken at every Kilometre mark along the Paholyothin Highway.

The data was checked and edited both manually and by a computer processing before storing in computer files for fast retrieval. To retrieve the information from the file, observation date and test site type are selected as the primary indices for each data set. Then each record may be retrieved in terms of its field identification.

ANNEX 4

PROGRAMMING PHASES OF THE RECOG PACKAGE

### PROGRAMMING PHASES OF THE RECOG PACKAGE

The RECOG package is divided into six phases which process the reformatting data from satellite computer compatible tapes. The performance of each phase is as described below :

- Phase One : This phase is used in conjunction with ground information to select training and test fields. It reads the spectral data from a RECOG reformatting CCT file and displays selected channels of desired area in form of a gray map or contour map. A histogram of data distribution in the specified area can also be plotted.
- Phase Two : This is a statistical processor for the training fields selected after studying Phase One output and ground information. The first statistical information printed consists of a correlation matrix, a mean vector, and a vector of a standard deviation for each selected field. Only the lower triangular portion of the covariance matrix and mean vector of each field are punched on cards which will be used for classification purpose in Phase Three, Phase Four and Phase Five. Histograms of the data distribution for selected channels in each field will be displayed to check if there is any fault selection pixel. The coincident class spectral plot will be printed as the last step to show the combination between the selected channels and the selected fields.
- Phase Three : The programme computes the divergence between each pair of fields for all possible combinations of a given member of channels in the covariance matrices and mean vectors. It will select the ten optimum wavelength bands which can best

discriminate between the fields of interest to the best of the divergence values.

Phase Four : This is a preclassification processor intended for an experimental categorial assignment regarding the decision schemes. Such decision schemes are level slicing, Euclidean distance and Gaussian Likelihood ratio. The programme reads the spectral data from a RECOG reformatting CCT file and reads the covariance matrices and mean vectors of the selected fields and then prints out the classified map according to the given decision scheme. Summary table, which consists of the numbers of pixels in each specified field, is also printed.

Phase Five : Only the Gaussian Likelihood ratio scheme will be used for this final classification. After the spectral data from a RECOG reformatting CCT file, covariance matrices and mean vectors of the selected fields are read by the programme, Phase Five will classify the data pixel according to the assigned values. Then it will write assigned category associated with its statistical value to the computer files for use in Phase Six.

Phase Six : The purpose of this phase is to thresh out the insignificant categorial assignment by using the Chi-Square table. Since Phase Five assigns every pixel of the given area to the category which gives the highest probability of belonging some misclassified pixels will be printed in the classified map. With the use of Chi-Square table in Phase Six, it will thresh out the pixels with low statistical significance by comparing to the number of channels used in the classification and the thresholding index.

\*\*\*\*\*